AD_____

Award Number: W81XWH-05-1-0391

TITLE: Maintaining Genome Stability: The Role of Helicases and Deaminases

PRINCIPAL INVESTIGATOR: XiaoJiang Chen

CONTRACTING ORGANIZATION: University of Southern California
Los Angeles, CA 90089-1147

REPORT DATE: July 2007

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
01-07-2007	Annual	1 Jul 2006 – 30 Jun 2007

4. TITLE AND SUBTITLE

Maintaining Genome Stability: The Role of Helicases and Deaminases

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-05-1-0391

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

XiaoJiang Chen

E-Mail: Xiaojiang.Chen@usc.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Southern California
Los Angeles, CA 90089-1147

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

No abstract provided.

15. SUBJECT TERMS
No subject terms provided

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			USAMRMC
U	U	U	UU	9	**19b. TELEPHONE NUMBER** *(include area code)*

TABLE OF CONTENTS:

Summary of Aims:

(Aim 1). Study the *in vitro* functions of MCM proteins from archaea and yeast cells using the genetically engineered protein constructs. In this aim, we will also extend our prior success in the X-ray structural studies of an N-terminal fragment of an archaea MCM by attempting to crystallize MCM proteins from yeast.

(Aim 2). Examine *in vivo* effects of helicase function and in particular MCM roles in maintaining genome integrity in response to damage. This aim will use existing and newly generated mutants, which can be achieved through genetic screening and site-directed mutagenesis based on the 3-dimensional structure of MCM, to investigate how MCMs contribution to genome stability during chemical damage.

(Aim 3). Express, purify and crystallize the proteins of deaminases. We will focus on AID and APOBEC3G to obtain purified deaminase proteins for the *in vitro* biochemical, functional, and structural studies.

(Aim 4). Examine the functions and substrate specificity of AID and identify other factors required for the coupling of deamination with other processes of DNA synthesis and RNA transcription. The experiments will be carried out in a cell free assay system, using already purified DNA replication/repair and RNA transcription proteins in our labs.

Progress for Aim 1 (*in vitro* function of MCM):

MCM regulator structure/function: While archaeal MCM is functional as a hexamer and double hexamer, the eukaryotic MCM is inactive as a hexemeric complex. A breakthrough finding is made by Michael Botchan's group, in which they showed that if a complex of GINS (tetrameric complex) binds to MCM in complex with CDC45, the MCM becomes active in unwinding DNA. Thus GINS complex is the "missing" link that regulates and activates MCM helicase activity, and thus DNA replication activity.

The GINS complex, which contains the four subunits, Sld5, Psf1, Psf2 and Psf3, is essential for both the initiation and progression of DNA replication in eukaryotes. GINS associates with the MCM2-7 complex and Cdc45 to activate the eukaryotic MCM helicase. It also appears to interact with and stimulate the polymerase activities of DNA polymerase ε and the DNA polymerase α-primase complex.

To further understand the functional role of GINS, in collaboration with Dr. Jerard Hurwitz's lab, we determined the crystal structure of the full-length human GINS hetero-tetramer. The four subunits each have a major domain composed of an α-helical bundle-like structure. With the exception of Psf1, other subunits have a small domain containing a three-stranded β-sheet core. Each full-length protein in the crystal has unstructured regions that are all located on the surface of GINS which are probably involved in its interaction with other replication factors. The four subunits contact each other mainly through α-helices to form a ring-like tetramer with a central pore. This pore is partially plugged by a 16-residue peptide from the Psf3 N-terminus which is unique to some eukaryotic Psf3 proteins and is not required for tetramer formation. Removal of this N-terminal 16 residues of Psf3 from the GINS tetramer increased the opening of the pore by 80%, suggesting a mechanism by which accessibility to the pore may be regulated. The structural data presented here indicate that the GINS tetramer is a highly stable complex with multiple flexible surface regions.

A model of how the human GINS tetrameric complex interacts and coordinates the activities of its binding partners is proposed in Fig. 6. In this model, we suggest a direct contact between Psf1 and Pol ε because an interaction between *S. cerevisiae* Psf1 and Dpb2 (the second largest subunit of Pol ε) was detected in a yeast two hybrid screen. Because archaeal Gins23 binds to the N terminus of MCM and an interaction of Psf3 with MCM has been detected in the yeast two hybrid system (S. Azuma and H. Masukata, unpublished data, cited in), our model depicts Psf3 in contact with MCM. We assume that Psf2 contacts the Pol α primase complex based on the report that archaeal Gins23 interacts with primase. The model also allows MCM, Cdc45, and GINS to contact each other based on the isolation of this complex from Drosophila eggs . Finally, we also suggest that GINS must interact with Pol ε and the Pol α-primase complex to coordinate leading and lagging strand synthesis, respectively.

Progress for Aim 2

To examine *in vivo* effects of helicase function and in particular MCM roles in maintaining genome integrity in response to damage. This aim will use existing and newly generated mutants in MCM proteins to investigate how MCMs contribution to genome stability during chemical damage. We have initiated a screen for *mcm* mutants that are functional for DNA replication, but defective in response to damage. The screen design is as follows. We mutagenized *in vitro* plasmids containing wild type *mcm2*$^{+}$, *mcm4*$^{+}$, and *mcm7*$^{+}$. These were transformed into strains containing temperature sensitive mutations in *mcm2, mcm4* and *mcm7*, and selected for viability at 36°C, at which temperature the chromosomal MCM is inactivated. Cell viability depends upon the plasmid-borne MCM. This ensures that the plasmid copy is functional for replication. Then, we screened the surviving transformants for sensitivity to hydroxyurea.

Our first screen was conducted under conditions that are lethal for checkpoint mutants such as Δ*rad3*. We did not observe any hydroxyurea-sensitive MCM clones. However, following the old genetics adage "absence of evidence is not evidence of absence", we are re-calibrating our screening conditions to use a concentration of HU where we can see a phenotype in Δ*cds1* checkpoint mutants. This strain is more resistant to HU, and our other experiments suggest that the Cds1 kinase may be involved directly in MCM function (see below), which makes it particularly relevant. We will also screen our transformants for sensitivity to other agents including MMS. Finally, we will screen existing *mcm* mutants using the same regimen.

To determine domains of MCMs responsible for interacting with other proteins, and investigate whether these interactions are mediated through single MCM subunits or the intact complex. As described in our preliminary data, we showed that the MCM proteins interact with the Rad51 (Rhp51) recombination protein. We have now shown that MCMs also interact with the Cds1 checkpoint kinase (human Chk2, budding yeast Rad53). These data are being submitted for publication. We hypothesize that MCMs

may be a substrate of this kinase. In particular, we observe that an *mcm4ts* mutant blocked in hydroxyurea, and then released to restrictive temperature, is competent to complete replication and go on to divide. In contrast, this mutant shifted to restrictive temperature without pre-incubation in HU blocks late in S phase with severe damage. This suggests that HU has a protective effect on the *mcm4ts* allele. Further support of this comes from the same experiment with an *mcm4-degron* mutant. This allele causes rapid degradation of the Mcm4 protein at the restrictive temperature. As expected, the *mcm4-degron* mutant has the same phenotype regardless of pre-incubation with HU: a lethal arrest in S phase. We hypothesize that pre-activation of Cds1 by the HU treatment protects the Mcm4ts protein. We are now testing to see whether the temperature sensitive protein interacts with Cds1 and is a Cds1 substrate, and whether this influences interaction with other checkpoint proteins (e.g., Rad17).

Progress for Aim 3

Structure-function study of deaminase enzymes

APOBEC-2 (Apo2) belongs to the Apolioprotein B (APOB) mRNA-editing enzyme catalytic polypeptide (APOBEC) family of cytidine deaminases found exclusively in verterbrates. APOBEC nucleic acid deaminases modify genes by deaminating cytosines in mRNA coding sequences and in ssDNA. Additionally, these enzymes can inhibit the replication of retroviruses, such as the human immunodeficiency virus (HIV) and hepatitis B virus (HBV), and retrotransposons.

The APOBEC family is composed of APOBEC-1 (Apo1), Apo2, AID, and APOBEC-3 (3A, 3B, 3C, 3F, 3G, and3H). Apo1, the first member to be characterized, deaminates $C^{6666} \rightarrow U$ in the APOB mRNA thereby creating a premature stop codon, which results in a truncated protein with a different function. Of the APOBEC3 subgroup, APOBEC-3B (A3B), APOBEC-3F (A3F) and APOBEC-3G (A3G) have two cytidine deaminase (CDA) domains and inhibit HIV-1 replication in the absence of the HIV viral infectivity factor (Vif). APOBEC3 proteins also shield the human genome from the deleterious action of endogenous retrotransposons: A3A, A3B, A3C and A3F inhibit LINE 1 and Alu retrotransposition.

AID and Apo2 have a single CDA homology domain and are phylogenetically the most ancient members of APOBEC family [1]. AID induces somatic hypermutation (SHM) and class switch recombination (CSR) in activated germinal center B cells. Point mutations in AID are responsible for an immunodeficiency disease, Hyper-IgM-2 (HIGM-2) syndrome, which is characterized by a deficiency in isotype-switching and high affinity antibody formation. Apo2, also known as ARCD-1, is ubiquitously expressed at low levels in both human and mouse and highly expressed in cardiac and skeletal muscle [15]. Pro-inflammatory cytokines, TNF-α and IL-1β, induce Apo2 expression in hepatocytes via the NF-kB response elements identified in the Apo2 promoter region. Apo2 and A3A are the only two APOBEC enzymes that have not yet been reported to deaminate cytosine in ssDNA or RNA *in vitro*. Apo2 is encapsulated into HIV-1 virions when co-expressed with Δvif HIV-1 DNA in 293T cells. However, studies fail to show that Apo2 and nearly all of the single domain APOBEC enzymes inhibit HIV-1 viral replication, except for A3C, which exhibits weak HIV-1 antiviral activity.

A comprehension of the molecular mechanisms of the APOBEC enzymes has been limited by the lack of 3-dimensional structures of APOBEC proteins that use large nucleic acid substrates. A major breakthrough was achieved through a collaboration of the Chen and Goodman laboratories that resulted in the first high-resolution crystal structure (2.5 Å) for an APOBEC protein, APOBEC-2, reported in Nature, Ref. (3). The structure is made up of an unusual rod-shaped tetramer that differs significantly from the square tetramer of the free nucleotide C deaminase. Although an enzymatic activity for APOBEC-2 has yet to be identified, its close homology with AID enabled us to use APOBEC-2 as a surrogate to predict how mutations in analogous regions might alter AID activity. This strategy proved to be remarkably effective, where we found that mutations that were predicted to influence dimer and tetramer interface interactions caused significant reductions in AID activity. In an analysis of "APOBEC2 structure – AID function by proxy", we examined 5 HIGM-2 human disease-associated

AID mutants that fail to make high-affinity antibodies. We found that one is mutated at the tetramerization interface, 4 are mutated on an exposed monomer surface and 1 mutation is located proximal to the catalytic active center. These data illustrate the importance of the multimeric interactions in the biochemical function of AID, and, more importantly, in its biological function.

Progress for Aim4

Initiation of Human Hypermutation by the APOBEC Family of Nucleic Acid Deaminases

Almost all mutations are either inconsequential or deleterious. And yet mutations play a central role in evolution, adaptation and fitness from microorganisms to humans. The "trick" is to select and then amplify cells carrying the very few beneficial mutations. Non-patogenic microorganisms, *Escherichia coli* and *Salmonella typhimurium* use error-prone DNA polymerase to enhance cell fitness in highly competitive environments, e.g., limited food resources, whereas their closely related pathogens shut off mismatch repair pathways resulting in "hypermutability". In humans, somatic hympermutation (SHM) of the variable region of immunoglobulin genes is responsible for generating high affinity antibodies.

We have made major progress in dissecting the biochemical basis for the initiation of hypermutation by two members of the APOBEC family of nucleic acid cytidine deaminases, AID (activation-induced cytidine deaminase) and APOBEC3G. AID is synthesized only in B cells, where tightly regulated synthesis occurs for about 2 hours. AID is required for somatic hypermutation and class-switch recombination of human immunoglobulin genes, acting during transcription of V-genes on the non-transcribed strand. APOBEC3G is encapsulated along with the genomic RNA in the HIV-1 virion. APOBEC3G plays an important role in blocking infection by the HIV-1 (AIDS virus) in non-permissive cells. Both enzymes deaminate cytidine to form uracil (C $\rightarrow$ U) on ssDNA substrates. Strict regulation of deamination is key, because unregulated deamination, caused by expression and subsequent action of deaminases in the wrong place at the wrong time, are likely to lead to serious disease, e.g., AID-initiated B-cell lymphomas, by causing excessive levels of mutation (typically a million-fold above normal mutation frequencies).

Our primary goal is to study the biochemical and physical biochemical mechanisms for the APOBEC family of "mutators". The close collaboration of Xiaojiang Chen's laboratory with our laboratory, fostered by the Army Grant support, has enabled us to marry biochemical, physical biochemical analysis with structural analysis, and we have succeeded in obtaining the first high-resolution crystal structure for any APOBEC enzyme. During the past grant period, we published the structure of APOBEC-2 (present in heart and skeletal tissue) in Nature.

Biochemical behavior of AID

We have shown previously that AID binds to randomly to ssDNA, irrespective of base composition, and catalyzes deamination of C favoring WRC (W = A/T, R = purine) hot spot motifs and disfavoring SYC (S = G/C, Y = pyrimidine) cold spot motifs. The enzyme scans processively along "naked" ssDNA and tracks processively on the non-transcribed strand of a moving transcription bubble. The deaminations occur in random clusters of between 4 and 10, usually containing at least one WRC triplet, interspersed with sparsely deaminated regions, i.e., regions with WRC motifs that are "ignored". The question is: what are the biochemical mechanisms? AID is believed to act as a dimer *in vivo* and its found to be phosphorylated at Ser38 in B cells.

The effects of phosphorylation on AID activity

During the previous grant period, because of its fundamental biological importance, we have placed a heavy emphasis on determining the effects of phosphorylation on AID activity and processivity. F. Alt and colleagues have established the importance of AID phosphorylation during SHM and CSR *in vivo*. We have compared phosphorylated wild type AID with dephosphorylated AID and AID mutants with mutations at predicted phosphorylation sites and have analyzed AID mutants with serine residues replaced by aspartic acid, which mimic phosphorylated AID (AID 'mimic mutants'). We have shown that AID purified from

baculovirus-infected insect cells is phosphorylated at several serine residues, including the most important Ser38. Alt and coworkers showed that phosphorylation of Ser38 is required for B-cell AID to initiate SHM and CSR *in vivo* and for activity on transcribed dsDNA, but not ssDNA *in vitro*. Although our data agree that Ser38 phosphorylation is needed for optimal AID activity, however, in contrast, we showed that both dephosphorylated AID and Ser38 to Ala mutant have a decreased specific activity on both ssDNA and transcribed dsDNA. The effect of phosphorylation appears principally to be electrostatic because replacement of Ser38 with Asp, "phosphorylation mimic mutation", yields fully active AID.

Much more importantly, we have constructed AID mutants that fail to do CSR and SHM in humans resulting in an inability to make high-affinity antibodies. These mutations, which are the source of the HIGM-2 syndrome, lead to an accumulation of low-affinity IgM antibodies. One of the HIGM AID mutants S38P cannot be phosphorylated at Ser38. Therefore, we were surprised to find that S38P had, for all intents and purposes, normal C deamination activity and hot and cold spot motif specificity. The clear import of this result is that the HIGM syndrome cannot simply be attributed to an inability of AID to catalyze C to U deamination *per se*. To understand the biochemical basis of the disease, our focus will turn toward the role of phosphorylation in substrate targeting. Once wild type AID or the S43P AID HIGM mutant binds to ssDNA, it catalyzes C to U deaminations avidly and with proper *in vivo* specificity. The question to be resolved is what is the mechanism of AID targeting to ssDNA? Perhaps even more to the point, how does AID target V-genes undergoing transcription while "ignoring" actively transcribed C genes, and what prevents AID from deaminating transcribed non-Ig genes? We are analyzing the ability of S38P to perform transcription-dependent deamination, and are concurrently determining the biochemical properties of other HIGM-2 AID mutants.

Biochemical behavior of APOBEC3G

Like AID, APOBEC3G scans ssDNA processively. However, unlike AID, deamination is directional with a strong 3' $\rightarrow$ 5' preference. Directional deamination is entirely unexpected because there is no obvious energy source, such as ATP hydrolysis. We have proposed that APOBEC3G slides and jumps along an ssDNA substrate in either direction, as does AID, but can only catalyze deamination of C when translocating 3' to 5'. It is known that the main target motif for APOBEC3G is CCC. We have observed that it is the bases immediately adjacent to the 3'-side of CCC that determine the probability of APOBEC3G-catalyzed deamination. The data suggest that APOBEC3G "sees" its target CCC motif principally from the 3'-side. The biological consequence of directional deamination is that it offers a plausible biochemical explanation for the well-established property of HIV-1, which has a much higher G $\rightarrow$ A mutational density toward the 3'-end of its RNA genome caused by having more deamination toward the 5'-end of cDNA by APOBEC3G. During the previous grant period, we published a paper in Nature Structural and Molecular Biology, Chelico et al., Reference (2), which describes this model. The NSMB paper was cited as a Research Highlight in Nature, 440, 2006, p. 1093, under the heading Molecular Biology, "Slide and jump". The paper was also featured in a Nature Structural and Molecular Biology News and Views, 440, 380-381 (2006), RS Harris & H Matsuo, entitled "Dancin' deaminase".

During the current grant period, we have delved into the directional APOBEC3G deamination mechanism by seeking conditions that either modulate or abolish directional deamination. Using linear single-stranded DNA substrates, we find that directional deamination is abolished in the absence of salt, and increases linearly with increasing salt concentrations until saturation occurs at high salt levels. Divalent cations (e.g., Mg^{2+}) enhance directional deamination at lower concentrations than monovalent cations (e.g., Na^{+}). Directional deamination is abolished on circular ssDNA substrates irrespective of salt, but directionality is restored when the circular DNA is made partially double-stranded by annealing with a short ssDNA complementary strand. We propose a working hypothesis in which a DNA end may provide a 3'-docking site at which APOBEC3G can bind and then wrap around the DNA in a 3' to 5' direction first by jumping and then by sliding. Favored deamination at hot spot motifs are located toward the 5'-end occurs because of the presence of a longer region of ssDNA (i.e., larger target size) in the 5'-direction.

Significance

The humoral response is able to recognize and attack an "unlimited" number of possible infectious agents by radically diversifying the inherited immunoglobulin (Ig) gene. First, the Ig gene is constructed from a variety of inherited gene segments by a Rag-mediated VDJ recombination event which joins together a variable (V), diversity (D) and joining (J) segments to make a functional Ig gene capable of expressing a low affinity antibody. In order to completely clear an infection, B cells undergo affinity maturation to produce antibodies with stronger affinity for antigen and with appropriate effector function. Affinity maturation involves two diversification events, somatic hypermutation and class switch recombination (SHM and CSR). Our principal long-term objective is to focus on SHM and to reconstitute the hypermutational process in an *in vitro* model system. Our initial objective is to explore the action of activation-induced cytidine deaminase (AID), an enzyme required for both SHM and CSR. We have shown that AID-catalyzed deamination of cytosine residues on ssDNA simulates three hallmark properties of SHM – mutational hot and cold spot sequences, broad clonal heterogeneity and transcriptional dependence. Active transcription is also a requirement for SHM, and our current data suggest that AID is able to target C $\rightarrow$ T mutations preferentially on the non-transcribed strand of a moving transcription bubble. An immediate priority of ours is to develop an *in vitro* human transcription system to investigate how AID is targeted to actively transcribed V- genes but not C genes. A biochemical understanding of AID, whose expression is confined to B cells, and the other APOBEC homologs, whose expression is diverse, should provide important new information on the potential risk factors, which we have referred to as "at risk nucleic acid sequences", which when acted upon inappropriately by APOBEC nucleic acid C deaminases can have serious deleterious consequences resulting from APOBEC-induced genomic instability, leading to cancer and perhaps even to neurodegenerative disease. Progress in understanding the molecular interactions governing SHM, which along with CSR is prerequisite to generating high affinity antibodies, would benefit substantially by the availability of an *in vitro* model system to study AID's role. We intend to provide this invaluable biochemical component.

Future Plans

Although we have progressed towards understanding the biochemical mechanisms responsible for initiating mutations by AID during SHM and by APOBEC3G during HIV-1 inactivation, an in-depth analysis will clearly require high-resolution structural analysis of wild type and mutant APOBEC family enzymes. During the next grant period, we will continue to work in close collaboration with Xiaojiang Chen to obtain crystals suitable for X-ray analysis. We have set forth on an ambitious agenda to achieve these goals. Drs. Chen and I have embarked on a joint effort with Dr. Raymond Stevens, Scripps, to try to crystallize all 10 APOBEC family members using Dr. Stevens' high throughput robotic strategy. The initial test of the high throughput strategy will be to see if crystals of APOBEC-2 can be obtained. Our main interest is to obtain structural data for AID and APOBEC3G. We will, however, deem this effort a success if we succeed in obtaining structural data for any of the other APOBEC proteins. We have recently purchased a multi-angle light scattering apparatus, which we will use to analyze the multimeric properties of APOBEC-2, AID and APOBEC3G. The goal is determine how structure alters function by performing site-directed mutagensis on each of the proteins, based on the structure of APOBEC-2, to introduce subtle perturbations at the dimer-dimer and tetramer interfaces, and also to introduce gross perturbations to abolish dimer and tetramer formation. The light scattering measurements will enable us to determine unambiguously whether each of the mutant proteins is in the form of a tetramer, dimer or monomer in solution.

Publications

1. Bransteitter, R., Sneeden, J., Allen, S., Pham, P., Goodman, M. F. First AID (Activation-induced Cytidine Deaminase) is Needed to Produce High-affinity Isotype-switched Antibodies. J. Biol. Chem. 281, 16833-16836 (2006).
2. Chelico, L., Pham, P., Calabrese, P., and Goodman, M. F. Jumping and Sliding, APOBEC3G DNA Deaminase Acts Processively 3' to 5' on Single-Stranded DNA. Nat. Struct. Mol. Biol. 13, 392-400 (2006).
3. Prochnaw, C., Bransteitter, R., Klein, M. G., Goodman, M. F., and Chen, X. S. APOBEC2 crystal structure and functional implications for AID. Nature 445, 447-451 (2007)